

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2002, 20:58:41 ; search time 1527.5 seconds  
(without alignments)  
16168.980 Million cell updates/sec

Title: US-10-025-514-15

Perfect score: 1525

Sequence: 1 tctagaccatggaagacct.....ccagtcaggcctagtgcac 1525

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estcpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_Other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	28.1	2478	11 AF130068	AF130068 Homo sapi
2	427.4	28.0	2571	11 AF113676	AF113676 Homo sapi
3	331.2	21.7	1392	11 AK002537	AK002537 Mus muscu
4	318.4	20.9	1296	11 AK004999	AK004999 Mus muscu
5	309.8	20.3	887	14 BQ643710	BQ643710 AGENCOURT
6	305.4	20.0	907	14 BQ648909	BQ648909 AGENCOURT

7	286.8	18.8	895	14	BQ653587	BQ653587 AGENCOURT
8	285.2	18.7	982	14	BQ646142	BQ646142 AGENCOURT
9	284.2	18.6	985	14	BM924019	BM924019 AGENCOURT
10	283.8	18.6	924	14	BQ958958	BQ958958 AGENCOURT
11	274.8	18.0	1194	14	BM924813	BM924813 AGENCOURT
12	273.4	17.9	907	14	BQ650189	BQ650189 AGENCOURT
13	270.4	17.7	959	14	BQ648524	BQ648524 AGENCOURT
14	268.6	17.6	891	14	BQ646948	BQ646948 AGENCOURT
15	268	17.6	795	12	BG567260	BG567260 602589741
c	266.4	17.5	833	14	BQ007663	BQ007663 UI-H-EIO-
16	264.8	17.4	991	14	BQ064738	BQ064738 AGENCOURT
17	263.6	17.3	687	10	AV649258	AV649258 AV649258
18	262	17.2	686	12	BG569467	BG569467 602589927
19	261	17.1	917	14	BQ650502	BQ650502 AGENCOURT
20	260.2	17.1	916	14	BQ644130	BQ644130 AGENCOURT
21	260.2	17.1	915	14	BQ644130	BQ644130 AGENCOURT
22	260.2	17.1	977	14	BQ654015	BQ654015 AGENCOURT
23	259.6	17.0	740	10	AV653778	AV653778 AV653778
24	259.6	17.0	944	14	BQ645533	BQ645533 AGENCOURT
c	259.6	17.0	1012	14	BQ064473	BQ064473 AGENCOURT
25	256	16.8	869	13	BI219058	BI219058 602938673
26	256	16.8	869	13	BI219058	BI219058 602938673
27	253	16.6	654	12	BG563537	BG563537 602581944
28	251.4	16.5	884	12	BG563537	BG563537 602581944
29	251.2	16.5	917	14	BQ650698	BQ650698 AGENCOURT
30	250.6	16.4	746	12	BG566873	BG566873 602589565
31	250.2	16.4	745	9	AI114643	AI114643 HA1237 Hu
32	249.8	16.4	1044	9	AI528087	AI528087 u193c09.Y
33	249.6	16.4	813	13	BI759000	BI759000 603042479
34	249.4	16.4	764	12	BF384318	BF384318 602046666
35	249	16.3	731	10	AV649144	AV649144 AV649144
c	249	16.3	813	10	BQ182052	BQ182052 UI-H-EUO-
36	249	16.3	813	10	BQ182052	BQ182052 UI-H-EUO-
37	246	16.1	804	12	BG618951	BG618951 602646361
c	244.4	16.0	780	9	AI110706	AI110706 HA0145 Hu
38	244.4	16.0	992	13	BI246204	BI246204 602958716
39	244.2	16.0	752	12	BG618001	BG618001 602644877
40	243.8	16.0	674	14	BM724546	BM724546 UI-E-E01-
41	243.2	15.9	759	12	BG533293	BG533293 601860666
42	243.2	15.9	749	9	AI526727	AI526727 u142b09.Y
43	243	15.9	749	9	AI526727	AI526727 u142b09.Y
44	242	15.9	629	12	BG545257	BG545257 602572457
45	241.2	15.8	780	12	BG617594	BG617594 602615086

## ALIGNMENTS

RESULT 1  
AF130068  
LOCUS  
DEFINITION Homo sapiens clone FLB8226 PRO2209 mRNA, complete cds.  
ACCESSION AF130068  
VERSION AF130068.1 GI:11493442  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2478)  
AUTHORS Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F.  
TITLE Functional prediction of the coding sequences of 75 new genes deduced by analysis of cDNA clones from human fetal liver  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2478)  
AUTHORS Zhang, C., Yu, Y., Zhang, S., Wei, H., Bi, J., Zhou, G., Dong, C., Zai, Y., Xu, W., Gao, F., Liu, M. and He, F.  
TITLE Direct Submission  
JOURNAL Submitted (23-FEB-1999) Department of Experimental Hematology, Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing, Beijing 100850, P. R. China  
FEATURES  
source  
Location/Qualifiers  
1. .2478  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="FLB8226"

```

/tissue_type="liver"
/dev_stage="fetus"
1082..2338
/notes="predicted protein of HQ2209"
/codon_start=1
/evidence=not_experimental
/product="PRO2209"
/protein_id="AAG35496.1"
/db_xref="GI:11493443"
/translation="WPSSVSMGILLLAGLCLVPLVSLAEDPQDAAOQKTDTSRHHDDH
PFFNKITPLAEFAESLRQLAHQSNSTNIPSPVSIATAFAMLSLKPADTHDEILE
GUNFLNTEIPFQAIHEGQELLRTLNQDSQILQITNGFLSEGLKLVDFLEDEKVK
LYHSEAFVFNFGDEEEDVDQVTVTKVPMKRLGNFNIHQKLSLSSWVLLMKYLGNAITF
FLPDEKLOHLENDTHDIITKFLNEDRRSASLHLPKLSITGYDLKSLVGLGKITK
VPSNGADLSGVTEEAAPLKLKSAVHKAVLTIDBKGEAAGAMFLEAIPAIPPEVKFNK
PFVFLMIEQNTSPLEMKVNPYTK"
BASE COUNT      650 a   627 c   597 g   604 t
ORIGIN
Query Match      28.1%; Score 429; DB 11; Length 2478;
Best Local Similarity 60.2%; Pred. No. 6.e-101;
Matches 711; Conservative 0; Mismatches 470; Indels 0; Gaps 0;
QY 12 GAAGACCTCAAGCGGACGCGCTCAAAAACCGACACCATCATCAGCACCAAGACCAT 71
DB 1154 GAGGATCCCGAGGAGATGCTGCCAGAGACAGATACATCCACCATGATCAGGATCAC 1213
QY 72 CCGACTTTTAAATAAATTTACTCAAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA 131
DB 1214 CCAACCTTCAACAGATCAACCCCACTGCTGAGTTCGCTTACGCTATACCGCCAG 1273
QY 132 TTAGCTCATCAAGATATCTTACTTAACATTTTTTTTAGTCTCTTTTCTATGCTCATGTCT 191
DB 1274 CTGCGACACCATGCAACAGCACCAATATCTTCTTCCCGAGTGAAGCATCGCTACAGCC 1333
QY 192 TTCGCCATGTGAGTTAGTACTAAAGCCGATACCATCATCAGAGATTTTGAAGTTTA 251
DB 1334 TTTGCAATGTCTCTCCCTGGGGACCAAGGCTGACACTCAGCATGAAATCCTGGAGGGCTG 1393
QY 252 AACTTTAATTTGACCGAAATCCAGAGACCCAAATTCAGGAGGTTTCAAGAGTTGTTG 311
DB 1394 AATTTCAACCTCAGGAGATTCGGAGGCTCAGATCCATCAAGGCTTCCAGGAACCTCTC 1453
QY 312 AGAAGTTGAATCAACCTGATCTCAATTTGCAATTAATCTACTGTTAGCGTTTATTTTG 371
DB 1454 CGTACCTCAACAGGACGACGAGCTCCAGCTCAGCTCAGGCAATGGCTGTTCTCTC 1513
QY 372 TCTGAAGCTTTAAATTTGTTGACAAATTCCTAGAAGAGCTCAAGAACTATATCATAGT 431
DB 1514 AGCGAGGCGCTGAAGCTAGTGGATGAATTTTGGAGGATGTTAAAGATGTTACCACTCA 1573
QY 432 GAGGCTTTTACCGTTAATTTTGGTGATCTACTGAGGAGCTAAAGACAAATTAATGATTAT 491
DB 1574 GAAGCTTCACTGTCACTTCGGGACACCCAGAGCCAGAAAGAAACAGATCAAGATTAC 1633
QY 492 GTTGAGAAGGCCACCGGTAAGATCGTTGACCTAGTTAAAGAAATTAAGATCGTGATACC 551
DB 1634 GTGGAGAAGGGTACTCAAGAGGAAAATTTGGGATTTTGGTCAAGGAGCTTGACAGAGACAC 1693
QY 552 GTCTTCGCACTAGTTAACTATATTTTTCAGGTAAGTGGGACGCTCTTTTCGAGGTT 611
DB 1694 GTTTTGTCTGGTGAATTTACATCTCTTTTAAAGGCAATGGGAGACCCCTTTGAAGTC 1753
QY 612 AAGATACTGAAGAGAGAGATTTTTCATGTTGATCAAGTTACTACTGTCAAAAGTTCCAAATG 671
DB 1754 AAGGACCGGAGAGAGAGACTTCCACGTGGACAGGTTGACCCGTTGAAGTGCCTATG 1813
QY 672 ATGAAAAGACTGGGTATGTTCAATATTCACATTCGCAAAAATTAAGTTCTTGGGTCTTA 731
DB 1814 ATGAAGCGTTAGGATGCTTTTAAATCCAGCATGTTAAGAGAGCTGTCCAGCTGGGTGCTG 1873
QY 732 TTAATGAAGTATTTAGGTAAACGCTACTCTATTTTTTTTTTACCAGAGGTAAGCTT 791

```

```

Db 1874 CTGATGAATFACCTGGCAATGCCACCGCATCTCTCTCTGCTGATGAGGGGAAACTA 1933
QY 792 CAACATTTTAGAATAGTGTGACTCATGACATTAATTAATTTTAGAGAGACGAGAT 851
Db 1934 CAGCACCTGGAAATGAATCACCACGATATCATCAAGTTCTCTGGAAATGAAGAC 1993
QY 852 CGTGTAGCGCTTCTCTGCACCTGCCAAAGTTAAGTATCATCCGCTACTTACGACTTAAA 911
Db 1994 AGAAGTCTGCCAGCTTACATTTACCCAAACTGTCCATTAATGGAACCTATGATCTCAAG 2053
QY 912 TCTGTTTTAGGCCAGTATAGTATTACCAAAAGTTTTTTTCTAACGGTCCGATTTGAGTGGT 971
Db 2054 AGCGTCTGGTCACTGGGCATCACTAAGTCTTCAGCAATGGGCTGACCTCTCCGGG 2113
QY 972 GTTACTGAAGAAGCTCCATTAATTAACCAATTCGTTTCTGTGATGATCGAGCAGAACACTAAA 1151
Db 2234 CCCCCGAGGTCAGCTTCAACAAACCCCTTGTCTTCTTAAATGATTGAACAAATACCAAG 2293
QY 1152 ACCCATTTGTTTATGGTAAAGTTGTCACACCACTCAGAA 1192
Db 2294 TCTCCCTCTTCATGAGAAAGTGGTGAATCCCAACCCAAA 2334

RESULT 2
AF113676 2571 bp mRNA linear HTC 08-MAY-2001
LOCUS Homo sapiens clone FLB2803 PRO0684 mRNA, complete cds.
DEFINITION AF113676
ACCESSION AF113676
VERSION AF113676.1 GI:6855600
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2571)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Ouyang,S., Luo,L., Wei,H., Zhou,G.,
Zhou,W., Bi,J., Zhang,Y., Liu,M. and He,F.
TITLE Functional prediction of the coding sequences of 32 new genes
deduced by analysis of cDNA clones from human fetal liver
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2571)
AUTHORS Zhang,C., Yu,Y., Zhang,S., Ouyang,S., Luo,L., Wei,H., Zhou,G.,
Zhou,W., Bi,J., Zhang,Y., Liu,M. and He,F.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1998) Department of Experimental Hematology,
Institute of Radiation Medicine, Beijing Taiping Road 27, Beijing
100850, P. R. China
FEATURES
Source Location/Qualifiers
1..2571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="FLB2803"
/tissue_type="liver"
/dev_stage="fetus"
/note="HQ0684"
515..819
/repeat_region
1109..2365
/note="predicted protein of HQ684"
/codon_start=1
/evidence=not_experimental
/product="PRO0684"
/protein_id="AAF29581.1"
/db_xref="GI:6855601"

```

/translation="MPSSVSMGILLLAGLCLVPSVLAEDPQDAQAOKTDRSHDDQH  
PTFNKITPLNFAEPSPYQRLAHQSNSTNIFSPVSIATAFAMLSLGTAKDFHDEILE  
GLNFPAITPEAIOIHGEOELLRTLNODSLOLQTTNGELFSLGKADFLDEYKK  
LYHSEAFVNGDTEFAKQINDYVEKGTQKINDVLYKELDRDVFALVNIFFKRW  
ERPVEKTEDEDFHVDQVTVKPMARLUGMFIHQHCKLUSWVLLMKYLGUNATAIF  
FLPDGKLOHLENEIHTDIIITKFLNEDRRSASLHPLKLSITGTYDLKSVLGOLGITK  
VFSNGADLSGVTEEAPLKLKSAHVHKAULTIDEKGTTEAAGAMFLEAIPMSIPPEVKFNK  
PFVFLMIQNTKSPFLMGKVVNPQK\*  
BASE COUNT 800 a 638 c 556 g 577 t  
ORIGIN

Query Match 28.0%; Score 427.4; DB 11; Length 2571;  
Best Local Similarity 60.1%; Pred. No. 1.7e-100;  
Matches 710; Conservative 0; Mismatches 471; Indels 0; Gaps 0;

QY 12 GAAGACCTCAAGGCGACCGCTCAAAACCGACACAGTCATCAGCACCAAGACCAT 71  
DB 1181 GAGGATCCCGGAGATGCTCCCGAAGACAGATACATCCACCATGATCAGGATCAC 1240

QY 72 CGGACTTTTAAATAAATTTACTCAAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA 131  
DB 1241 CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTCCGCTTTCAGCCCATACCGCCAG 1300

QY 132 TTAGCTCATCAAGTAATTTCTACTAACATTTTATTTAGTCTGTTTCTATTGCCACTGCT 191  
DB 1301 CTGGCACACAGTCCCAACAGCACAATATCTTCTTCTCCAGTGCAGATCGCTACAGCC 1360

QY 192 TTGCGCATGTTGAGTTAGTACTAAGCCGATACCCATGACGAGATTTTGAAGGTTTA 251  
DB 1361 TTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCAGATGAATCTGAGGCGCTG 1420

QY 252 AACTTTAATTTGACCGAATCCCGAAGCCCAATTTACGAGGGTTTTCAGAGAGTTGTTG 311  
DB 1421 AATTCAACCTCAGGAGATTCGGAGGCTCAGATCCATGAAGGTTTCAGGAACCTCCTC 1480

QY 312 AGAAGCTTGAATCAACCTGATTTCTCAATTCGAATTAATCTGTTGTAACGGTTTATTTTGG 371  
DB 1481 CGTACCTTCAACAGCCAGACAGCAGCTCCAGCTGACCCAGCGCAATGGCGTGTCTCTC 1540

QY 372 TCTGAAGGTTTAAATTTGTTGACAAATTCCTAGAGAGCTCAAGAACTATATCATAGT 431  
DB 1541 ACGGAGGCTGAGCTAGTGGATAGTTTGGAGGATGTTAAAGTGTGTACACATCA 1600

QY 432 GAGGCTTTTACCGTTAATTTTGGTGATCTAGGAGAGCTAAAGCAAAATTAATGATTAT 491  
DB 1601 GAAGCCTTCACTGTCAACTTCGGGACACCGGAGGCGCAAGACATCAAGATTTAC 1660

QY 492 GTTGAGAAAGGACCCAGGTTAGATCGTTGACCTAGTTTAAAGAAATTAAGATCGTGATACC 551  
DB 1661 GTGGAGAAGGGTACTCAAGGGGAAATTTGGATTTTGGTCAAGGAGCTTGACAGAGACACA 1720

QY 552 GTCTTCGGCACTAGTTAACTATATTTTTCAGGAGTAAGTGGAGAGCTTCCTTCGAGGTT 611  
DB 1721 GTTTTGTCTGTTGAATTTACATCTCTTAAAGCAATGGGAGAGACCTTTTGAAGTC 1780

QY 612 AAAGTACTGAAGAGAAATTTTCATGTTGATCAAGTTACTACTGTCAAAAGTTCCCAATG 671  
DB 1781 AAGGACCCGAGGAGAGGACTTCCACGTGGACAGGTTGACACCGCTGAAGTGCTCATG 1840

QY 672 ATGAAGAGCTGGGTATGTTCAATATTCACATTCGCAAAATTAAGTTCTTTGGTCTTTA 731  
DB 1841 ATGAAGGTTTATGAGGATGTTTAAACATCCAGCACTGTAAGAAGCTGTCCAGCTGGGTG 1900

QY 732 TTAATGAAGTATTTAGTTAAGCTACTGCTATTTTTCAGGAGCAAGGTAAGCTT 791  
DB 1901 CTGATGAATACCTGGGCAATGCCACCCCATCTCTCTCTGCTGATGAGGGGAACTA 1960

QY 792 CAACATTTAGAAATGAGTTGACTCATGACATTAATTTTACTAAATTTTATAGAACAGGAT 851  
DB 1961 CAGCACCTGGAAATGAACCTCACCCAGCATATCATCACAAGTTCTCTGGAAATGAAGAC 2020

QY 852 CTCTAGGCTTCTCTGCACTGCGCAAGTTAGTATCATCCGGTACTTACGACTTAAA 911

Db 2021 AGAAGGTCGCCAGCTTACATTTTACCACAACTGTCCATTACTGTGGAACCTATGATCTCAAG 2080

QY 912 TCTGTTTATAGCCAGTTAGTATTACCAAGTTTTTTTCTTAAGGTCGCCGATTTGAGTGGT 971  
Db 2081 AGCGTCTCGGTCAACTGGGATCACTAAGGTTCTTCAGCAATGGGCTGACCTCTCCGGG 2140

QY 972 GTTACTGAAGAAGCTCCCATTTAAATTTAGTAAAGCTGTTCACAAAGCGGTCTTAACTATT 1031  
Db 2141 GTCACAGAGGAGGACCCCTGAAGCTCTCCNAGGCCGTGCATAGGCTGTGCTGACCATC 2200

QY 1032 GATGAAAGAGGTACCGAGCGCGCGCTATGTTCTTCTGGAAGCTATTTCCAATGAGCAT 1091  
Db 2201 GACGAGAAAGGACTGAAGCTGCTGGGCCATGTTTTTAGAGGCCATACCATCTCTCAT 2260

QY 1092 CCACGAGAGTTAAATTTAATAAACCATTCGTTTTTCTGATGATCGAGCAGAACACTAAA 1151  
Db 2261 CCCCCGAGGTCAAGTTTCAACAAACCCCTTTGTTCTTCTTAATGATGAACAAAATACCAAG 2320

QY 1152 AGCCATTTTATGGGTAGGTTGTCAACCCCAACTCAGAA 1192  
Db 2321 TCTCCCTCTTCATGGGANAAGTGTGAAATCCCAACCAAAA 2361

RESULT 3  
AK002537  
LOCUS  
DEFINITION  
Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610011G14:serine protease inhibitor 1-4, full insert sequence.  
ACCESSION  
AK002537  
VERSION  
AK002537.1 GI:12832592  
KEYWORDS  
HPC; CAP trapper.  
SOURCE  
Mus musculus (strain:C57BL/6J) adult male kidney cDNA to mRNA, clone:0610011G14.  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE  
9279253  
PUBMED  
10349636

REFERENCE  
AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)  
JOURNAL  
2049374  
MEDLINE  
11042159  
PUBMED

REFERENCE  
AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, I., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Tameda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE  
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer  
Genome Res. 10 (11), 1757-1771 (2000)  
JOURNAL  
20530913  
MEDLINE  
11076861  
PUBMED

REFERENCE  
AUTHORS  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,

Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M., Wadner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seva, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y.

**TITLE** Functional annotation of a full-length mouse cDNA collection

**JOURNAL** Nature 409 (6821), 685-690 (2001)

**MEDLINE** 21085660

**PUBMED** 11217851

**REFERENCE**

**AUTHORS**

5 (bases 1 to 1392)

Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hiraoka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamanaka, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.

**TITLE** Direct Submission

Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@res.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

**COMMENT** Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGGCGCGCACTCGAGTTTCTTTTCTTNN 3'], cDNA was prepared by using trihalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGGATCCAGAGTCATTAATTAATTAATTAACCCGCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: SstI; 3' end: XhoI. Host: SOLR.

**FEATURES** Location/Qualifiers

1. .1392  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="FANTOM.DB:0610011G14"  
/db\_xref="MGD:MGI:1896776"  
/db\_xref="taxon:10090"  
/clone="0610011G14"  
/sex="male"  
/tissue\_type="kidney"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="adult"

1. .1392

/gene="Sp11-4"

68. .1309

/gene="Sp11-4"

/note="data source:MGD, source key:MGI:891968,

evidence:ISS

putative

serine protease inhibitor 1-4"

/codon\_start=1

/protein\_id="BAB22173.1"

/db\_xref="GI:12832593"

/db\_xref="MGD:MGI:891968"

/translation="MTPSISWLSLLALLGCLCPVSLAEVDVQETDTSQKQSPASHEI  
ATNLGDFALRLRYRELHVQSNSTNFFSPVSIATAPFAMLSLGSKGDTHQILEGLQFNL  
TQTSEADTHFQHLQTLNPDSLELQSTGNLGFVNDLUKLVKFELEAKNHQFAEV  
FSVFAESEEAKVINDVEKGTQKIVEAVKLDQDQTFALANYILFKGKWKQFPDP  
ENTPEAEHVDSEKTVKVPMTLSGLMDVHSCMSLWSWVLMIDYAGNTAVFLPDDG  
KMHLEOPLNKLISQFLNRRSRDAQIHIPRLSIGNYNLKTLMSPLGTRIFNNGA  
LDSGITEENAPLKLSKAVHKAVLIDETGTEAAATVLOVATYSMPPIVRDHPFLFI  
IFEHTQSPFIVGVKAVVDPTHK"

BASE COUNT 364 a 393 c 324 g 311 t

ORIGIN

Query Match 21.7%; Score 331.2; DB 11; Length 1392;

Best Local Similarity 56.3%; Pred. No. 1.9e-75;

Matches 662; Conservative 0; Mismatches 508; Indels 6; Gaps 2;

QY 20 TCAAGCGGACGGCGCTCAAAAAACCGACACCATCATCAGCACCAGCAAGCCATCCGACATT 79

DB 133 TCTGGCTGAGGATGTTTCAGGAGACACACACCTCCCGAAGAGGATCAGTCC---CCAGCCTC 189

QY 80 TAATAAAATTACTCCTCAAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACATTTAGCTCA 139

DB 190 CCATGAGATCGTACAAACCTGGGACATTTGCCCTCAGACTATACAGGAGGCTGGTCCA 249

QY 140 TCAAGTAATTTCTACTAACATTTTCTTTTGTAGTCTCTGTTCTTATTTGCCACTGCTTTGCCCAT 199

DB 250 TCAGTCCACACTTCCACACATCTTCTTCCCCAGTGACCATGCCACAGCCTTTGCTAT 309

QY 200 GTTGAGTTTGTAGTACTTAACCCGATACCATCAGCAGATTTTGAAGGTTTAAACTTTAA 259

DB 310 GCTCTCCTTAGGAGCAAGGGTGACACTCACACGAGATCCTAGAGGGCTCGAGTTCAA 369

QY 260 TTTGACGGAATCCCAAGACCCCAATTTACAGAGGTTTTCAGAGGTTTCTTGAGAACTTT 319

DB 370 CCTCAGCAAAACATCGGAGGCTGACATCACAAAGTCTTCCCAACCTCCTCCCAACCT 429

QY 320 GAATCAACCTGATTCCTCAATTTGCAATTAAGTCTGTAACGGTTTATTTTGTCTGAAGG 379

DB 430 CAACACACAGACAGTGTGAGCTGACGCTGACACAGCAATGGCTCTTTGTCAACAATCA 489

QY 380 TTTAAATTTGTTGACAAATTCCTAGAACGCTGAAGAACTATATCATGATGAGGCTTT 439

DB 490 CTTGAAGCTGTTGAGAAAGTTTCTGGAAGAGGCAAGAACCATTTATCAGGCAAGATCTT 549

QY 440 TACCGTTAATTTGGTGATCTCAGGAAGCTTAAAGCAAAATTAATGATTGTTTGAGAA 499

DB 550 CTCTGCAACTTTGCAAGGTCAGAGAGGAGGCAAGAAAGTATTAAATGATTGTTGGAGAA 609

QY 500 AGCACCCAGGGTAAGATCGTCCCTAGTCTTAAAGAAATTAGATCGTGATACCGTCTTCGC 559

DB 610 GGAACCCAGGAAGAAATAGTTGAGCGGGTGAAAGAACTGGACCAAGACACAGTTTCGC 669

QY 560 ACTAGTTAATATATTTTTCAGGGTAAGTGGGAGCGTCTCTTCGAGGTTAAAGATAC 619

DB 670 CTTGGGCAATTTACATTTCTTTTAAAGCAAAATGGAACAGCCATTCGATCCTCGAGAAC 729

QY 620 TGAAGAGGAAGATTTTCATGTTTCAATCAAGTTTACTGTTCTGTCCTCAAGTATGAAAG 679

DB 730 TGAGGAGAGCTGAGTTTCCATTTGAGGAGGCTCCACAGTGAAGGTCCTCCATGATGACC 789

QY 680 ACTGGGTATCTTCAATATTTCAACATTTGCAAAATTAAGTCTTTGGGTCTTTATTAATGA 739

DB 790 CTGGGCAATTTGATGTGATCATTTGCAGCATGCTGTCCAGCTGGTGTCTGCTGATGGA 849

QY 740 GTATTAGGTAAGCGTACTGCTATTTTATTTTACAGAGGAAGGTAAGCTTCAACATTT 799

DB 850 TTACGAGGCAACACCACTGCT 909

QY 800 AGAGATGAGTTGACTCATGACATTTTACTAAATTTTATAGAGAACGAGGATCGTCTGAG 859







using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

BASE COUNT	242 a	224 c	232 g	187 t	2 others	
ORIGIN						
Query Match	20.3%; Score 309.8; DB 14; Length 887;					
Best Local Similarity	60.0%; Pred. No. 6.5e-70;					
Matches	532; Conservative 0; Mismatches 356; Indels 1; Gaps 1;					
QY	195	GCATGTTGAGTTAGTACTAAGCCGATACCCATACGAGATTTTGAAGGTTTAAAC	254			
Db	1	GCAATGCTCTCCCTGGGACCAAGGCTGACACTCAGATGAATCTGGAGGCGCTGAAT	60			
QY	255	TTTAATTTGACCGAATCCAGAGCCCAATTCAGGAGGTTTCAAGAGTTTGAGA	314			
Db	61	TTCACCTCAGGAGATCCGGAGGCTCAGATCCATGAAGGTTCCAGAACTCCCTCGT	120			
QY	315	ACTTTGAATCAACCTGATTCTCAATTCGAATTAATTAATTAATTAATTAATTAAT	374			
Db	121	ACCTCAACACGACACAGCCAGCTCCAGCTGACCCACCGCANAGGCTGTTCCTCAGC	180			
QY	375	GAAGGTTTAAATTTGTTGACAAATTCCTAGAGAGCTCAAGAACTATATCATAGTGAG	434			
Db	181	GAGGGCTGAAGCTAGTGATAAGTTTGGAGGATGTTTAAAGTTGTACCACTCAGAA	240			
QY	435	GCTTTTACCGTTAATTTTGTGATGACTGAGGAGCTTAAAGCAATTAATGATTATGTT	494			
Db	241	GCCTTCACCTGTCACTTCGGGACACCGAAGAGGCGCAAGAAACAGATCAACGATTACG	300			
QY	495	GAGAAAGGACCCAGGTTAAGATCGTTGACCTAGTTTAAAGAAATTAAGTCTGATACCCG	554			
Db	301	GAGAAGGTTACTCAAGGGAATTTGGATTGTTCAAGGACTTGCAGAGACACAGT	360			
QY	555	TTCCGCTAGTTAACTATATTTTTCAGGGAAGTGGGAGCGCTTTCAGGTTTAA	614			
Db	361	TTTGCTCTGGTGAATTTACATCTCTTTTAAAGCAATAGGAGAGACCCCTTTGAAGTCA	420			
QY	615	GATACGAAAGAGGAGATTTTCATGTTGATCAAGTTACTACTGTCMAAGTTTCCAATGATG	674			
Db	421	GACACGAGGAGAGGACTTCCAGCTGGACAGGTGACCCCGTGAAGTGCTATGATG	480			
QY	675	AAAGACTGGGTATGTTCAATATTCACATTCGAAATTAAGTTCTTGGCTCTTATTA	734			
Db	481	AAGCGTTTAGGATGTTTAAATCCAGCACTGTAAGAAGCTGCCAGCTGGGTGCTGCTG	540			
QY	735	ATGAGTATTTAGGTAACGCTACTGCTATTTTTCACGAGAGGAGGTAAGCTTCAA	794			
Db	541	ATGAATACCTGGGCAATGCCACCGCACTCTTCTTCCCTGCTGATGAGGGAACTACAG	600			
QY	795	CATTTAGAGATGAGTTGACTCATGACATATTACTATAATTTTATAGAGACGAGGATCGT	854			
Db	601	CACCTGGAATATGAACTACCCACCATATCATCCACAGTTCCCTGGAAATGAAGACAGA	660			
QY	855	CGTAGGCTTCTCTGCACCTGCCAAAGTTAAGTATCACCGTACTTAGGACTTAAATCT	914			
Db	661	AGGTCTGCCAGCTTACATTTACCCAACTGCTCCATTACTGGAACCTATGATCTGAAGC	720			
QY	915	GTTTTAGGCCAGTTAGTATTACCAAGTTTTCCTACGGTGGCGATTTAGTGGGTGT	974			
Db	721	GTCTGNGTCAACTGGGCATCACTAAGTCTTTCAGCAATGGGCTCTCCCGGNGTC	780			
QY	975	ACTGAAGAGCTCCATTAATTTAGTAAAGCTGTTCAAGCGCTCTTAACCTATTGAT	1034			
Db	781	ACAGAGGAGCACCTTGAGCTCTCCAGGC-GTGCAATAGGCTGTGCTGACCATCGAC	839			
QY	1035	GAAAGGTPACCGAGCGCGCGCTATGTTCTCGAAGCTATTCC	1081			
Db	840	GAGAAAGGAGTGAAGTCTGGGGCCATGTTTTTAGAGGCCATACC	886			
RESULT 6						
BQ648909						

## LOCUS

DEFINITION BQ648909 907 bp mRNA linear EST 15-JUL-2002

5', mRNA sequence. AGNCOURT\_8351269 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6286839

ACCESSION BQ648909

VERSION BQ648909.1 GI:21773081

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 907)

TITLE NTH-MGC http://mgs.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: CGAP (Stanford)

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM2488 row: e column: 16

High quality sequence stop: 721.

FEATURES

Location/Qualifiers

1..907

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:6286839"

/clone\_lib="NIH\_MGC\_100"

/tissue\_type="hepatocellular carcinoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: liver; Vector: pORF7; Site:1: XhoI; Site:2:

EcoRI; CDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(G). Size-selected &gt;500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

BASE COUNT 246 a 228 c 239 g 192 t 2 others

ORIGIN

Query Match 20.0%; Score 305.4; DB 14; Length 907;

Best Local Similarity 59.8%; Pred. No. 9.2e-69;

Matches 530; Conservative 0; Mismatches 356; Indels 1; Gaps 1;

QY 217 AAGCGATACCCATGACGAGATTTTGAAGGTTTAACTTTAATTTGACCGAATCCAG 276

Db 15 AGGCTGACACTCAGATGAATCTGGAGGCTTGAATTTTCAACCTCAGGAGATCCGG 74

QY 277 AAGCCCAATTCACGAGGTTTCAAGAGTTGTTGAGAACTTTGAATCAACCTGATTC 336

Db 75 AGGCTCAGATCCATGAAGGCTTCCAGGAATCTCCCTGACCTCAACCCAGACAGCC 134

QY 337 AATTGCAATTAACCTACTGGTAACGGTTTATTTTGTCTGAAGGTTTAAAAATTGTTGACA 396

Db 135 AGCTCCAGCTGACCCACCGCAATGGCTGTCTCTCAGCAGGCGCTGAAGCTAGTGGATA 194

QY 397 AATTCTTAGAAGAGCTCAAGAACTATATCATAGTAGGCTTTTACCGTTAATTTGGTG 456

Db 195 AGTTTGGAGGATGTTAAAAAGTTGTACCACTCAGAAAGCTTCACTGTCAACTCGGG 254

QY 457 ATACTGAGGAGCTAAAAAGCAATTAATGATTATGTTGAGAAAGGACCCAGGTTAAGA 516

Db 255 ACACGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAGGGTACTCAAGGGAAA 314

QY 517 TCGTTGACCTAGTTTAAAGAAATTAGATCGTATACCGCTTTCGCACCTAGTTTAACATATTT 576

Db 315 TTGTGGATTTTGTCAAGGAGCTTGACAGACACAGTGTGTCTGCTGTAATTACATCT 374

QY	577	TTTTCAAGGTAAGTGGACACGCTCTTTCGAGGTAAAGATACTGAAGAGGAAGATTTTC	636
Db	375	TCITTAAGGCAATGGAGAGACCTTTCAAGTCAAGGACCGGAGGAGGACTTCC	434
QY	637	ATGTTGATCAAGTTACTACTGTCACAAAGTCCCAATGATGAAAGACATGGGTATGTTCAATA	696
Db	435	ACGTGGACCAAGTGACCAAGTGAAGGTGCTATGATGAAGCGTTTAGGCATGTTTAAACA	494
QY	697	TTCAACATTCGCAAAATTAAGTCTCTGGCTCTATTAAATGAAGTATTTAGGTAACGCTA	756
Db	495	TCCAGCACTGTAAGAAGCTGTCCAGCTGGGTGCTGCTGATGAATACCTGGGCAATGCCA	554
QY	757	CTGCTATTTTTTTTACCAGACGAAGCTAAGCTTCAACATTTAGAGAATGAGTTGACTC	816
Db	555	CGGCATCTTCTCTGCTGATGAGGGAACACTACAGCACCTGGAAATGAACCTCACCC	614
QY	817	ATGACATATTACTAAATTTTAGAGAACGAGGATCGTCTAGCGCTTCTCTGCACCTGC	876
Db	615	ACGATATCATCACCAGTTCTCTGAAATGAAGACAGAAGGTCTGCCAGCTTACATTAC	674
QY	877	CAAAAGTTAAGTATCACCGGTACTTACCACCTTAAATCTGTTTTAGGCCAGTTAGGTATTA	936
Db	675	CCAACTGCTCAATCTTGGAACTATGATCTGAAGACGCTCCTGGTCACTGGGCATCA	734
QY	937	CCAAAGTTTTTCTAAGCGTCCGATTTGAGTGGTGTACTGAAGAGCTTCCATTAAT	996
Db	735	CTAAGTCTTCAGCAATGGGCTGACCTCTCGGGGTACAGAGGAGGACCCCTGAAGC	794
QY	997	TGAGTAAAGCTGTTCAAAAGCGCTCTTAACCTATGATGAAGGGTACCAGGCGCCG	1056
Db	795	TCTCCAGGCG-GTGCAATAAGGCTGTGTGACCATCGACGAGAAAGGACTGAAGCTGCTG	853
QY	1057	CGCTATGCTCTCGAAGCTATTCCTCAATGAGCATTCACACAGAAAGTT	1103
Db	854	GGGCAATGTTTTAGAGGCAATACCATGCTAATCCCCCGAGGT	900
RESULT 7			
LOCUS	B0653587	895 bp	mRNA
DEFINITION	AGENCOURT_8349591 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6284550	linear	EST 15-JUL-2002
ACCESSION	B0653587	5', mRNA sequence.	
VERSION	B0653587.1	GI:21777759	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 895)		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-r@mail.nih.gov		
	Tissue Procurement: CGAP (Stanford)		
	cDNA Library Preparation: Rubin Laboratory		
	CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	Plate: LCM2482 row: f column: 07		
	High quality sequence stop: 708.		
	Location/Qualifiers		
	1. 895		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6284550"		
	/clone_lib="NIH_MGC_100"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2:		
FEATURES			
source			

RESULT 8

EcORI; cDNA made by oligo-dT priming. Directionally cloned into EcORI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 227 a 248 c 230 g 190 t

Query Match	18.8%	Score 286.8	DB 14	Length 895
Best Local Similarity	60.1%	Pred. No. 6.6e-64		
Matches 477	Conservative 0	Mismatches 317	Indels 0	Gaps 0
QY	12	GAAGACCCCTCAAGGGGACGCCGCTCAAAAACCCGACACCATGTCATCAGACCAAGACCAT	71	
Db	95	GAGGATCCCCGAGAGATGCTGCCAGAGAAGACAGATATCCACCATGATCAGGATCAC	154	
QY	72	CGGACTTTTAATAAATTTACTCCAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA	131	
Db	155	CCAACCTTCAACAAGATCACCCCAACCTGGCTGAGTTGCGCTTCAGGCTATACGCCGAC	214	
QY	132	TTAGCTCATCAAAAGTAATTTCTACTAACATTTTTTTTAGTCTCTGTTTCTATTGCCACTGT	191	
Db	215	CTGGCACACCATGTCACAGACCAATATCTTCTCTCCCAAGTGGCATCGCTACAGCC	274	
QY	192	TTGCGCATCTTGAGTTTAGGTACTAAAGCCGATACCCATGACGAGATTTTGAAGAGTTTA	251	
Db	275	TTTGCAATGCTCTCCCTGGGGACCAAGGCTGACACTCACGATGAATCTCTGAGGCGCTG	334	
QY	252	ACATTTAAATTTGACCGAAATCCAGAGCCCAATTCACGAGGTTTTCACAGTTGTTG	311	
Db	335	AAATTTCAACCTCACGAGATTCGGAGGCTCAGATCCATGAGGCTTCAGGAACTCTCT	394	
QY	312	AGAACTTTGAATCAACCTGATTTCTCAATTTGCAATTAACCTACTGTAACCGTTTATTTTG	371	
Db	395	CGTACCTCAACCCAGCAGACAGCCAGCTCCAGTACCCAGCCCAATGGCTGTCTCTC	454	
QY	372	TCGTGAGGTTTAAATTTGTTGACAAATTTCTAGAGAGCTCAAGAACTATATATAGTAT	431	
Db	455	AGCGAGGCTGAAAGCTAGTGGATAAGTTTGGAGGATGTTAAAAAGTTGTACCACTCA	514	
QY	432	GAGGCTTTTACCGTTAATTTTGGTGATGATCTGAGGAAGCTTAAAAAGCAAAATTAATGATT	491	
Db	515	GAAGCCTTCTCTGCTCACTTCGCGGACCGAGAGGCCCAAGAAACAGATCAACGATTAC	574	
QY	492	GTTGAGAAAGGCCACCGAGGTAAAGTCGTTGACCTAGTTAAAGATTAGTCTGTATAC	551	
Db	575	GTGGAGAGGGTACTCAAGGGAAATTTGGATTTTGGTCAAGGAGCTTGACAGAGACACA	634	
QY	552	GTCTTCGCCTAGTTAACTATATTTTTTCAAGGGTAAGTGGGAAGCTCTTTCGAGGTT	611	
Db	635	GTMTTGTCTCTGTTGAATTACATCTCTTTAAAGGCAATGGGAGAGACCCCTTGAAGTC	694	
QY	612	AAAGATCTGAAGAGGAAGATTTTTCATGTTGATCAAGTTTACTACTGTCAAAGTTTCAATG	671	
Db	695	AAGCACCCGAGGAGGAGCTTCCAGCTGGACAGCTGACCCAGGAGGTCCTATG	754	
QY	672	ATGAAAGACTGGGTATGTTTCAATATTCACATTTGCAAAATTAAGTTCTTGGGTCTTA	731	
Db	755	ATGGAGCGTTTAGGCATGTTTAAACATCCAGCACTGTGAAGAGCTGCCAGCTGGGCTG	814	
QY	732	TTAATGAAGTATTTAGGTAAAGCTACTGCTATTTTTTTTTTACAGAGGAAGTAAGCTT	791	
Db	815	CTGATGAATTAACCTGGGGCATGCCCGCCATCTTCTTCTGCTGGATCAGGGAACCTA	874	
QY	792	CAACATTTAGAGAA	805	
Db	875	CAGCACTGGGAAA	888	



BQ646142  
LOCUS BQ646142 982 bp mRNA linear EST 15-JUL-2002  
DEFINITION AGENCOURT\_8492569 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6296341  
5', mRNA sequence.  
ACCESSION BQ646142  
VERSION BQ646142.1 GI:21770314  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 982)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: CGAP (Stanford)  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM2504 row: a column: 14  
High quality sequence stop: 647.  
Location/Qualifiers  
1. 982  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6296341"  
/clone\_lib="NIH\_MGC\_100"  
/tissue\_type="hepatocellular carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: liver; Vector: pCRB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."  
BASE COUNT 267 a 259 c 238 g 217 t 1 others  
ORIGIN  
Query Match 18.7%; Score 285.2; DB 14; Length 982;  
Best Local Similarity 59.8%; Pred. No. 1.8e-63;  
Matches 496; Conservative 0; Mismatches 333; Indels 1; Gaps 1;  
QY 303 GAGTTGTTGAGAACTTTGAATCAACCTGATTCTCAATTGCAATTAACTACTGGTAACGGT 362  
DB 1 GAACCTCTCCGTACCTCAACAGCAGACAGCCAGCTCCAGCTGACCCCGGCNATGGC 60  
QY 363 TTATTTTGTCTGAGAGTTTAAATTTGGTTGACAAATTCCTAGAGAGCTCAAGAACTA 422  
DB 61 CTGTTCTCAGGAGGGGCTGAAGCTAGTGGATAAGTTTGGAGGATGTTAAAAAGTTG 120  
QY 423 TATCATAGTAGGCTTTTACCGTTAATTTTGGTGATCTAGGAGCTTAAAGCAATTT 482  
DB 121 TACCACTCAGAGCCCTTCACTGCTCAACTTCGGGGACCCGAGAGGCCAAGAAACAGATC 180  
QY 483 AATGATTATGTTGAGAAAGGCCAGCCAGGTAAGATCGTTGACCTAGTTTAAAGAAATTAGAT 542  
DB 181 AACGATTAGCTGAGAGAGGTTACTCAAGGAAATTTGGTATGGTCAAGAGCTTGAC 240  
QY 543 CTGTATACCGCTTCGCACTAGTTAACTATATATTTTTCAGAGGTAAGTGGGAACGTCCT 602  
DB 241 AGAGACACAGTTTTTGCCTCTGGTGAATTTACATCTCTTTTAAAGGCAATTTGGAGAGCC 300  
QY 603 TTCGAGGTTAAAGATCTAGAGAGAAATTTTCATGTTGATCAAGTTACTACTGTCAA 662  
DB 301 TTTGAAGTCAAGGACCCGAGGAGAGGACTTCCACGTGGACAGGTGACCACCGTGAAG 360

QY 663 GTTCCATGATGAAGAAGACTGGGTATGTTCAATATTCAACATTTCAAAAAAATTAAGTTCT 722  
DB 361 GTGCTATGATGAAGCGGTTTAGGATGTTTAACTATCCAGCATGTAAGAAGCTCTCCAGC 420  
QY 723 TGGGTCTTATTAAATGAAGTATTTAGGTAACGCTACTGCTATTTTATTTTATTTTACCAAGCA 782  
DB 421 TGGGTCTGCTGATGAATACCTGGGCATGCGCAATGCCACCGCATCTTCTTCTGCTGCTGATGAG 480  
QY 783 GGTAAAGCTTCAACATTTAGAGAAATGAGTGTGACTCATGACATTAATTAATAATTTTAAAG 842  
DB 481 GGGAAACTACAGCACCTGGAAAAATGAATCAACCCAGCATATCATCAACCAAGTTCTCCTGGA 540  
QY 843 AACGAGGATCTGCTAGCGCTTCTGCACTGCAACCTGCAAGTAAAGTATCAACCGGTACTTAC 902  
DB 541 AATGAAGACAGAAGGTCTGCCAGCTTACATTTACCCAACTGTCCATTTACTTGGAACTAT 600  
QY 903 GACTTAAATCTGTTTTAGGCGAGTTAGGTATTAACCAAGT-TTTTCTAACGGTGCGCA 961  
DB 601 GATCTGAAGAGCGTCTCTGGGTCACTGGGCATCACTAAGGTCTTTCAGCAATGGGGCTGA 660  
QY 962 TTTGAGTGGTGTACTGAAGAAGCTCCATTAATAATGAGTAAAGCTGTTTCAAAAGCGGT 1021  
DB 661 CCTCTCGGGGTCAAGAGGAGGACCCCTGAAGCTCTCCAAGCGCTGCATTAAGGCTGT 720  
QY 1022 CTTAACTATTGATCAAAAAGGTACCGAGGCGCGCGGCTATGTTCTGGAAGCTATTCC 1081  
DB 721 GCTGACCATCGACGAGAAAGGAGTGAAGCTGTGGGGCCATGTTTGAAGACCATACC 780  
QY 1082 AATGAGCATTCACACGAGAAGTTAAATTAATAAACCATTCGTTTTTCTGA 1131  
DB 781 CATGCTATCCCTCCCGGAGGAGGAGTTCAACCAACCCCTTGTCTTCTTA 830  
RESULT 9  
BM924019  
LOCUS AGENCOURT\_6709923 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5760183  
DEFINITION 5', mRNA sequence.  
ACCESSION BM924019  
VERSION BM924019.1 GI:19374398  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 985)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM12806 row: m column: 16  
High quality sequence stop: 707.  
Location/Qualifiers  
1. 985  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5760183"  
/clone\_lib="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector:  
pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
source anonymous pool of 3 colons, age 26 yo male, 49 yo  
female, 71 yo male colon; 46 yo male kidney, and pool of 2  
stomachs, 62 yo male and 70 yo female. Library is

oligo-dr primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.4 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 023. Note: this is a NIH\_MGC Library."

BASE COUNT	258 a	270 c	242 g	213 t	2 others
ORIGIN					
Query Match	18.6%;	Score 284.2;	DB 14;	Length 985;	
Best Local Similarity	60.3%;	Pred. No. 3.2e-63;			
Matches 503;	Conservative 0;	Mismatches 329;	Indels 2;	Gaps 2;	
QY	12	GAAGACCTCTAAGGCGAGCGCGCTCAAAAAACCGACACAGTATCATCCAGCAAGACCAT	71		
Db	92	GAGGATCCCGAGGAGATGTCGCCAGAGACAGATATCCACCATGATCAGGATCAC	151		
QY	72	CCGACTTTTAATAAATTTACTCCAAATTTAGCCGAATTTGCTTTTCTTTGTATAGACAA	131		
Db	152	CCAACCTTTCAACAAGATCACCCCAACCTGGCTGAGTTCGGCTTCAGCCTATACCGGCAG	211		
QY	132	TTAGCTCATCAAGTAATTTCTACTAACATTTTTTTTAGTCTCTGTTCTATGCCCACGTCT	191		
Db	212	CTGGCACACCAAGTCCACAGACCAATATCTCTCTCCCGAGTGGATCGCTACAGCC	271		
QY	192	TTCCGCATGTTGAGTTTAGGTACTAAAGCGCATACCCATGACGAGATTTTGAAGGTTTA	251		
Db	272	TTTGCAATGCTCTCCCTGGGACCAAGCTGACACTCAGATGAATCCCTGGAGGCGCTG	331		
QY	252	AACTTTAATTTGACCGAATCCAGAGACCCAAATTTACAGAGGTTTTCAGAGATTTGTTG	311		
Db	332	AATTTCAACCTCACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGAGAACTCCTC	391		
QY	312	AGAACTTTGAATCACTGATCTCAATTTGCAATTAATCTACTGTAGCGTTTATTTTGG	371		
Db	392	CGTACCTTCAACAGCCAGACAGACAGCTCCAGCTGACCCCGCAATGCGCTTCTCCTC	451		
QY	372	TCTGAAGGTTTAAATTTGGTGACAAATTTCTAGAAGACGTCAAGAACTATATCATAGT	431		
Db	452	ACCGAGGCGCTCAAGCTAGTGATAAGTTTTTTGGAGGATGTTTAAAGGTTTCTACCACTCA	511		
QY	432	GAGGCTTTTACGTTAATTTTGGTGATCTACTGAGNAGCTTAAAGCAAAATTAATGATAT	491		
Db	512	GAAGCCTTCTACTGTCAACTTCGGGACACCCGAAGAGGCCCAAGAAACAGATCAACGATTAC	571		
QY	492	GTGAGAAAGGCACCCAGGTAAGATCTGTACCTAGTTTAAAGAAATAGATCGTGATACC	551		
Db	572	GTGGAGAGGGTACTCAAGGGAATTTGGATTTGGTTCAGGAGCTTGACAGAGACACA	631		
QY	552	GCTTTGCGCACTAGTTAACTATATTTTTTCAAGGGTAAGTGGGAACGTCCTTTTCGAGGTT	611		
Db	632	GTTTTTGTCTGCTGAATTTACATCTCTTTTAAAGGCAATGGGAGACCCCTTTGAAGTC	691		
QY	612	AAAGATCTGAAGAGAGAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTTCCAAATG	671		
Db	692	AAAGACCCCGAGAGAGAGACTTCCACCTGGACAGGTTGACCACTGGAAGTGGCTATG	751		
QY	672	ATGAAAGACTCGGTATGTTCAATATTCACATTCGAAATTAAGTTCTTTGGGTCTTTA	731		
Db	752	ATGAGCGTTTAGGCATGTTTAAACATCCAGCACTGTAGAGCTGTTCCAGCTGGGTGCTG	811		
QY	732	TTAATTAAGTATTAGGTAACGCTACTGCTAT-TTTTTTTTTTACCAGACGAGGTAAGCT	790		
Db	812	CTGATGAATACCTGGCAATCCACCGCCATCTTCTTCNTGGCCTGATGAGGGGAACT	871		
QY	791	TCACAAATTTAGAGATGAGTGTGACTCATGACATTAATTAATTTTATAGAGAA	844		
Db	872	ACAGGCACTGGAAATGAAGTCTCAC-CAGGATATCATCAAGGTTTCTCTGGAAA	924		
RESULT 10					
BQ958958					
LOCUS	924 bp	mRNA	linear	EST 21-AUG-2002	

DEFINITION AGENCOURT\_10034701 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6483305  
5', mRNA sequence.  
ACCESSION BQ958958  
VERSION BQ958958.1 GI:22374436  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-re@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM2667 row: k column: 18  
High quality sequence stop: 586.  
Location/Qualifiers  
1. .924  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6483305"  
/clone\_lib="NIH\_MGC\_40"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pOTB7; Site1: XhoI;  
Site2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

#### FEATURES

Source

BASE COUNT 251 a 253 c 226 g 193 t 1 others  
ORIGIN

Query Match 18.6%; Score 283.8; DB 14; Length 924;  
Best Local Similarity 58.4%; Pred. No. 4e-63;  
Matches 513; Conservative 0; Mismatches 363; Indels 2; Gaps 1;

QY	84	AAAATTAATCTACTAACATTTTTAGTCCTGTTTCTATTGCGCACTGCTTTGCGCATGTTG	143
Db	2	AGATACCCCAACCTGGCTGAGTTCCGCTTACGCTATACCGCCAGCTGGCAGACACAG	61
QY	144	AGTAATTTCTACTAACATTTTTTTTAGTCCTGTTTCTATTGCGCACTGCTTTGCGCATGTTG	203
Db	62	TCCAACAGCACCATACTCTTCTCCCAAGTGGATCGCTACAGCCTTTTGCATGCTC	121
QY	204	AGTTTAGGTACTAAAGCCGATACCCATGACGAGATTTTGAAGGTTTAACTTTAATTG	263
Db	122	TCCCTGGGGACCAAGGCTGACACTCAGATGAATCCTGGAGGGCTGAAATTTCAACCTC	181
QY	264	ACGAAATCCAGAACCCCAATTCAGAGGGTTTTCAAGAGTTGTTGAGAACTTTGAAT	323
Db	182	ACGGAGATTCGGAGGCTCAGATCCATGAAGGCTTCAGGAATCTCCGTACCTCAAC	241
QY	324	CAACCTGATTCTCAATTTGCAATTAACCTACTGTTTAACTGTTTATTTTGTCTGAAGTTTA	383
Db	242	CAGCCAGACCCAGCTCCAGCTGACCCCGCAATGGCCTGTTCTCCAGCGAGGCGCTG	301
QY	384	AAATTTGGTTGACAAATTTCTTAGAAGAGCTCAGAACTATATCATAGTAGGCTTTTACC	443
Db	302	AAGCTAGTGGATAAGTTTTTGGAGGATGTTTAAAAAGTTGTACCACTCAGAAAGCTTTCACT	361
QY	444	GTTAATTTTGGTGATCTAGGAGCTAAAAAGCAAAATTAATGATTTATGTTGAAAGGC	503

Db 362 GTCAACTCGGGACACCGAAGAGGCCAAGAAACAGATCAACGATTACGTGGAGAAGGCT 421  
QY 504 ACCAGGGTAAGATCGTTCACCTAGTTAAAGAAATAGATCGTGATACCGCTTCGCCACTA 563  
Db 422 ACTCAAGGGAAATGTGGATTGTGCAAGGAGCTTTGACAGAGACACAGTTCCTGCTG 481  
QY 564 GTTAACATATATTTTTCAGGGAAGTGGGAAACGTCCTTCGAGGTTAAAGATACCTGAA 623  
Db 482 GTGAATACATCTCTTTAAAGGCAATGGAGAGACCCCTTTGAAGTCAAGGACACCGAG 541  
QY 624 GAGGAAGATTTTCATGTTGATCAAGTACTACTGTCACAAATGTCACAAATGATGAAGAACTG 683  
Db 542 GAAGAGGACTTCCAGCTGGACGAGGCGACCCGTCGAGGTCCTATGATGAAGCTTTA 601  
QY 684 GGTATCTTCAATATTCACCAATTCGAAATTAAGTTCCTGGGTCTTATTAATGAAGTAT 743  
Db 602 GGCATGTTTAACATCCAGCAGCTGTAGAAGCTGTCCAGCTGGGTGCTGATGAATATAC 661  
QY 744 TTAGTAAACGCTACTGCTATTTTTCACGAGGTAAGCTTCAACATTTAGAG 803  
Db 662 CTGGCAATGCCACGCCATCTTCTCGCTGATGAGGGAACACTACAGCACCTGGAA 721  
QY 804 AATGAGTTCACATGACATTAATTAATAATT--TTTGAAGAACGAGGATCGTCGACG 861  
Db 722 AATGAACCTCACCCAGATATCATCAGCAGTTCCTGGAAATGAAGACAGAAAGTCCTG 781  
QY 862 CTCTCTGACCTGCCAAAGTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTAG 921  
Db 782 CCACCTTAGATTACCAAACTGTCATTTCTGGAACATATGATGGAAGAGCGCTCTG 841  
QY 922 GCCAGTTAGTATTACCAAAAGTTTTTCTTAACGGTGCC 959  
Db 842 GTCAACTGGGCATCACTAAGGCTTTCAGCAATGGGGGC 879

RESULT 11  
BM924813 1194 bp mRNA linear EST 12-MAR-2002  
LOCUS AGENCOURT\_6653921 NIH\_MGC\_116 Homo sapiens cDNA clone IMAGE:5761267  
DEFINITION 5', mRNA sequence.  
ACCESSION BM924813  
VERSION BM924813.1 GI:19375192  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE NIH-MGC <http://mgi.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LMML)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LMML at:  
<http://image.llnl.gov>  
Plate: LLML2809 row: j column: 20  
High quality sequence stop: 662.  
Location/Qualifiers  
1..1194  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5761267"  
/clone\_11b="NIH\_MGC\_116"  
/lab\_host="DH10B"  
/note="Organ: pooled colon, kidney, stomach; Vector:  
pCMV-SPORT6; Site:1; Not1; Site:2; EcoRV (destroyed); RNA  
source anonymous pool of 3 colons, age 26 yo male, 49 yo  
female, 71 yo male colon; 46 yo male kidney, and pool of 2

FEATURES  
source

stomachs, 62 yo male and 70 yo female. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.4 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by C.  
Gruber (Invitrogen). Research Genetics tracking code  
023. Note: this is a NIH\_MGC Library."

BASE COUNT 321 a 303 c 328 g 242 t  
ORIGIN

Query Match 18.0%; Score 274.8; DB 14; Length 1194;  
Best Local Similarity 61.2%; Pred. No. 9.7e-61;  
Matches 444; Conservative 0; Mismatches 282; Indels 0; Gaps 0;  
QY 448 ATTTTGTGATCTAGGAGAGCTAAAAGCAAAATTAATGATTAATGTTGAGAAAGGCACCC 507  
Db 7 AATTTCCGGGATCCGAAGAGGCCAAGAAACAGATCAACGATTAAGTTCGGAGAAAGGCTACT 66  
QY 508 AGGTAAGATCTGTTGACCTAGTTAAAGANNTAGATCTGATACCGTCTTCGCACACTAGTTA 567  
Db 67 AAGGAAATTTGTGGATTGTCAGGAGCTTTGACAGAGACACAGTTCCTTTCCTCTGCTGA 126  
QY 568 ACTATATTTTTCAGGGTAAGTGGGAAGCTCTTTTCGAGGTTTAAAGATCTGAAGAGG 627  
Db 127 ATTACATCTTCTTTAAAGGCAATGGCAGAGACCCCTTTGAAGTCAAGCAGACCCGAGG 186  
QY 628 AAGATTTTCATGTTGATCAAGTTACTACTGTCAAAGTTTCCAAATGATGATAAAGACTTGGGTA 687  
Db 187 AGGACTTCCAGCTGGAGGCTGACACCGTGAAGGTGCTATGATGAAGCGTTTAGGCA 246  
QY 688 TGTTCATATATCAACATTTGCAAAATTAAGTTCCTCGGCTTATTAATCAAGTATTTAG 747  
Db 247 TGTTCATATCAACATCTGTAAGAGCTGTCCAGCTGGGTGCTGCTGATGAATACCTCG 306  
QY 748 GTAAAGCTACTGCTATTTTTCAGCAGCAAGGTAAGCTTCAACATTTTAGAAGATG 807  
Db 307 GCAATGCCACCGCCATCTTCTCTGCTGATGAGGGGAACTACAGCACCTGGAAATG 366  
QY 808 AGTTGACTCATGACATTAATTAATTTTTCAGAGAGGAGATCGTCGAGCGCTTCTC 867  
Db 367 AACTCACCCACGATATCATCAAGTTCCTGGAATAATGAAGACAGAAAGCTCTGCCAGCT 426  
QY 868 TGCACCTGCCAAAGTTAAGTATCACCGTACTTACGACTTAAATCTGTTTTAGGCCAGT 927  
Db 427 TACATTTACCAAACTCTCCATTAAGTATGATCTGAAGAGCTCTCGGGTCAAC 486  
QY 928 TAGTATTAACAAAGTTTTCCTAACGGTCCGATTTTGTGTTTACTGAAAGAGCTC 987  
Db 487 TGGGCATCACTAAGTCTTCAGCAATGGGCTGACCTCTCCGGGTACACAGAGAGGCAC 546  
QY 988 CATTAATTTGAGTAAAGCTGTTCAAAAGCCGCTTAACTATGATGAAGAGGGTACCG 1047  
Db 547 CCCTGAAGCTCTCCAGGCGCTGATAAGGCTGTGCTGACCATCGAGCAAGAGGACTG 606  
QY 1048 AGGCGCGCGGCTATGTTCTCTGGAAGCTATTCATATGAGCATTCACCAAGAGTTAAAT 1107  
Db 607 AAGCTGCTGGGCCATGCTTTTTCAGAGCCATACCCATGCTATCCCCCGGAGGCTCAAGT 666  
QY 1108 TTAATAAACCATCTGTTTTCCTGATGATGATGATGATGATGATGATGATGATGATGATG 1167  
Db 667 TCAACAACCCCTTTGCTCTTAATGATGATGATGATGATGATGATGATGATGATGATG 736  
QY 1168 GTAAGG 1173  
Db 727 GAAAG 732

RESULT 12  
BQ650189 907 bp mRNA linear EST 15-JUL-2002  
LOCUS AGENCOURT\_8298326 NIH\_MGC\_100 Homo sapiens cDNA clone IMAGE:6269613  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ650189

```
VERSION B0650189.1 GI:21774361
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2443 row: g column: 22
High quality sequence stop: 650.
FEATURES
    source
        1..907
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="NIH_MGC_100"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library."
BASE COUNT 245 a 237 c 226 g 198 t 1 others
ORIGIN
Query Match 17.98; Score 273.4; DB 14; Length 907;
Best Local Similarity 59.48; Pred. No. 2.le-60;
Matches 498; Conservative 0; Mismatches 337; Indels 3; Gaps 2;
QY 336 CAATTCGAATTAACCTACTGGAACGGTTTATTTTGTCTGAAGGTTTAAATTTGGTTGAC 395
DB 7 CAGCTCCAGCTGACCCAGCGCAATGGCTGTCTCAGCGAGGGCCTGAAGCTAGTGGAT 66
QY 396 AAATTCCTAGAGAGCTCAAGAACTATATCATAGTGAGGCTTTTACGGTTAATTTTGGT 455
DB 67 AAGTTTTTGGAGGATGTTAAAGATTGTACCACTCAGAAGCCCTTCACTGTCAACTTCGGG 126
QY 456 GATACCTAGGAGCTPAAAGCAATTAATCATATGTTTGAAGAGGCCACCGGGTAAG 515
DB 127 GACACCGAAGGCCAAGAAACAGATCAACCATTCGTTGGAGAGGGTACTCAAGGGAA 186
QY 516 ATCTGTGACCTAGTTAAAGAAATTAGATCGTGTATACCGTCTTCGCAGTATGTTAATATT 575
DB 187 ATGTGGATTGTTGTCAGAGGCTTGACAGAGACACAGTTTTTGTCTGTGTAATTACATC 246
QY 576 TTTTCAAGGGTAAGTGGAGAGCTCCTTTCGAGGTTTAAAGTACTGAAGAGAGAGATTTT 635
DB 247 TTTCTTAAAGGCAATGGGAGAGACCCCTTTGAAGTCAAGAGCACCGAGAGAGGACTTC 306
QY 636 CATGTTGATCAAGTTACTGTCACAGTCCCAATGATGAAGAAAGACTGGGTATGTTCAAT 695
DB 307 CAGGTGGACAGGTGACCAACCGGTGAAGGTGCTATGATGAAGCGTTTACGATGTTTAC 366
QY 696 ATTCACATTTGCAAAATTAAGTTCTTTGGGCTTTTATTAATCAAGTATTTAGGTAACGCT 755
DB 367 ATCCAGCACTGTAAAGAGCTGTCCAGCTGGGTGCTGCTGATGAATACTCTGGCAATGCC 426
```

```
QY 756 ACTGCTATTTTTTTTACCAGACGAAGGTAGCTTCAACATTTTAGAGAATGAGTTGACT 815
DB 427 ACCGCCATCTTCTCTCCCTGATGAGGGGAACTACAGCACCTGGAAATGAACACTACC 486
QY 816 CATGACATTTACTAAATTTTTTAGAGAACGAGGATCGTCGTAGCGCTTCTCTGCACTG 875
DB 487 CAGGATATCATCACCAGTTCTCTGGAATGAAGACAGAGAGGTCTGCCAGCTTACATTTA 546
QY 876 CCAAGAGTTAAGTATCACCGGTACTTACGACTTAAATCTGTTTGTAGGCCAGTTAGTATT 935
DB 547 CCAAACTCTCCATTTACTGGAACCTATGATCTGAAGAGCGTCTGGGTCAACTGGGCATC 606
QY 936 ACCAAGTTTTTCTAAACGTCGCGATTTGAGTGTGTACTGAAGAAGAGCTCCATTTAAA 995
DB 607 ACTAAGGTTCTCAGCAATGGGCTGACCTCTCCGNGTCCAGAGAGAGGCCCTCGAAG 666
QY 996 TTGAGTAAAGCTGTTTACAAAGCCGCTTCTTAATGATGAAAGGGTACCGAGCCGCC 1055
DB 667 CTCTCAAGGCCGTGCATAAGGCTGTGCTGACCATCGACGAGAGAAAGG-ACTGAAACTGCT 725
QY 1056 GGGCTATGTTCTTGAAGCTATTTCAATGAGCATTCCACGAGAGTTAAATTTAA--TA 1113
DB 726 GGAGCCATGTTTTTAGAGGGCTTACCCATGCTATCCCGCCCGAGTCAAGGTTCCACCA 785
QY 1114 AACCATTCGTTTTTCTGATGATCGAGCAGACACATAAAAGCCCATTTGTTTATGGGTAA 1171
DB 786 ACCCTTTGCTTTCTTTAATGATGGACCAATACCAAGTCTCCCTCTCTCCGGGGAA 843
```

```
RESULT 13
LOCUS B0648524
DEFINITION B0648524 959 bp mRNA linear EST 15-JUL-2002
5', mRNA sequence.
ACCESSION B0648524
VERSION B0648524.1 GI:21772696
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 959)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2444 row: d column: 16
High quality sequence stop: 674.
FEATURES
    Location/Qualifiers
        1..959
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="NIH_MGC_100"
            /tissue_type="hepatocellular carcinoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: liver; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally cloned
            into EcoRI/XhoI sites using the following 5' adaptor:
            GGACGAG(G). Size-selected >500bp for average insert size
            1.8kb. Library constructed by Ling Hong in the laboratory
            of Gerald M. Rubin (University of California, Berkeley)
            using ZAP-cDNA synthesis kit (Stratagene) and Superscript
            II RT (Life Technologies). Note: this is a NIH_MGC
            Library."
```

ACCESSION BQ646948  
VERSION BQ646948.1 GI:21771120  
KEYWORDS EST.  
SOURCE human.

